图书情报工作
第63卷第4期 2019年2月

基于模块度的话题发现及网民情感波动研究
——以新浪微博"中美间贸易摩擦"话题为例

■ 张海涛[1,2] 刘雅姝[1] 张枭慧[1] 宋拓[1]
[1] 吉林大学管理学院 长春 130022 [2] 吉林大学信息资源研究中心 长春 130022

摘要:［目的/意义］探索热点事件评论网络中话题社群及网民的情感波动，掌握舆情事件发展过程，对于整体把握热点事件的发展方向，做好新时期网络舆论的引导工作具有重大意义。［方法/过程］以复杂网络理论为基础，基于评论词语间的共现关系构建基于事件发展的子事件网络，通过社群发现算法来识别子事件评论网络中的话题社群，将情感词依据情感词典赋予情感分类属性，基于事件的演化过程动态地跟踪网民意见以及情感波动。［结果/结论］研究结果表明，评论网络群落发现以及变异系数方法可以有效地衡量网民话题讨论的规模与集中程度；评论网络中赋予情感词节点情感分类属性方法可以体现事件演化过程中网民的情感变化；舆论衍生话题对事件的舆情发展有持续性影响；网民话题讨论内容对于事件演化具有一定程度上的前瞻性。

关键词: 复杂网络 社群发现 话题发现 情感分析
分类号: G25
DOI: 10.13266/j.issn.0252-3116.2019.04.001

1 引言

随着互联网等各种新媒体手段的飞速发展和手机、移动设备的广泛应用以及社会化媒体为特征的媒介化时代的到来，网络日益成为反映社会舆情的主要载体和窗口。网络舆情[1]是指由个人以及各种社会群体构成的公众，在一定的历史阶段和社会空间内，就自己关心或与自身利益紧密相关的各种公共事务通过互联网表达和传播的各种不同情绪、态度和意见交错的总和。网民通过微博针对某一事件发表自己的观点和见解过程中会产生大量具有情感的文本信息，通过观点性信息进行话题发现与情感分析研究，可以了解网民的情感走势以及对于事件的整体看法和意见。网络舆情事件具有传播速度快、隐蔽性强、不易控制等特点，因此动态跟踪网民对舆情事件话题讨论内容以及情感的变化，将其作为重要的民意收集平台，对于整体把握热点事件的发展方向，充分利用网络媒体的特点做好新时期网络舆论的引导工作具有重大意义。

2 研究现状

2.1 微博话题发现

关于微博主题挖掘问题的相关研究主要集中于计算机领域，以改进经典聚类算法提高话题发现有效等问题为主。Y. Chen等[2]学者开发了一个增量聚类框架来检测新的主题，并采用一系列内容和时间特征来及时发现热门主题；G. Stilo等[3]基于相关时间序列的相似性提出了一种用于在微博中对词语进行聚类的新方法；M. Hu等[4]从用户评价中挖掘用户的观点；李亚星等[5]采用改进Single-Pass算法，提出基于实时共现网络的话题发现模型；宋莉娜等[6]提出了SOM聚类的微博话题发现方法，以改善传统文本聚类的不足，进而有效发现话题。此外，图书情报学领域也对此类问题积极展开研究，唐小波[7-8]等基于先验概率的潜在语义分析模型LDA对微博主题进行挖掘，提出用依存句法分析来改进传统文本相似矩阵，以提高聚类准确性。王正成等[9]利用LDA主题模型挖掘出特定话题中的各个主题，体现特定话题的舆情走向并针对主题挖掘意见领袖。肖璐等[10]基于句子成分的相似度计算方

作者简介：张海涛(ORCID:0000-0002-9421-8187)，教授，博士生导师；刘雅姝(ORCID:0000-0003-0742-5901)，博士研究生，通讯作者，E-mail:1911670944@qq.com；张枭慧(ORCID:0000-0003-0020-3126)，硕士研究生；宋拓(ORCID:0000-0003-1282-1861)，博士研究生。
收稿日期:2018-05-16 修回日期:2018-09-26 本文起止页码:6-14 本文责任编辑:易飞

法来计算文本之间的相似度,构建文本相似矩阵,进行聚类分析,找到微博热点主题。

2.2 微博情感分析

情感分析也被称为意见挖掘,是对带有情感色彩的主观性文本进行分析、处理、归纳和推理的过程[11]。情感分析主要有机器学习的方法和基于情感词典的方法两大类。分析大量用户发布的微博在很多领域都非常关键,挖掘用户情感具有广泛的应用价值,吸引了国内外许多研究人员的关注。国外学者对于网民情感变化的研究主要集中在 Twitter 平台上,B. Pang 等[12－13]利用人工标注训练语料,采用贝叶斯、最大熵等方法分析电影评论倾向性。J. Bollen 等[14]对发布在微博平台上的推文进行情绪分析,计算每天时间轴上六维情绪向量,进而推测大规模的情绪分析。X. Zou 等[15]提出了一种结合社会情景和话题情景来分析微博情绪的新方法,引入话题上下文来模拟微博之间的语义关系。L. Zhang 等[16]提出了一种新的实体层面的情绪 Twitter 的分析方法。由于中文微博与英文微博在产品设计上存在不同,中文和英文在语言表达习惯上也有很大差异,因此与英文微博的情感分析相比,中文微博的情感分析研究工作还不是很成熟。一部分学者提出关于特定研究问题的情感分类方法,唐晓波等[17]提出一种基于特征本体的产品评论情感分析法,该方法利用构建特征本体对特征词进行分类,通过计算情感词与特征的搭配权重来识别隐式特征,构建领域情感词典和微博表情符号词典。杨亮等[18]提出情感分布语言模型,通过分析相邻时间段情感分布语言模型间的差异,实现对热点事件的发现。黄卫东等[19]提出一种基于概率潜在语义分析(PLSA)的网络舆情话题情感分析方法,利用 PLSA 模型对不同时段上的网络舆情话题进行子话题提取和情感词表构建。也有学者使用复杂网络理论进行情感分类研究,张向阳等[20]根据消费者在线评论,构建有向网络,从评论网络的拓扑性质出发建立在线评论情感倾向分类模型。杨锋等[21]基于随机网络以及情感词表提出了一种基于评论序列最短覆盖路径的情绪倾向性分类方法。

综上所述,学者们基于不同情感分类方法构建情感词表、利用情感词表预测用户特定行为或进行博文内容分析以及对优化各类聚类算法以提高话题发现方面的研究较多,将复杂网络思想引入网民情感波动与话题意见挖掘的研究较少。有鉴于此,本研究选取新浪微博热门话题"中美间贸易摩擦"子事件中热门博文下更能体现网友态度的评论数据作为数据源,将评论中词语作为网络节点,根据情感词典赋予情感词节点与之相对应的情感分类属性,依据词语共现关系确定网络中的边,进而构建评论网络,并根据 Louvain 算法[22],划分评论网络群落,动态跟踪热点事件网民话题意见和情感波动变化情况,进而显示出舆情事件完整的演进状况。

3 基础理论与研究框架

3.1 基础理论

3.1.1 网络社群划分　复杂网络是研究解决复杂系统问题的一种重要方法。复杂网络的理论及实证研究对于生物学传染病的传播研究、蛋白质表达网络的研究、社会学市场扩展网络研究、物理学中广泛现象研究、计算机科学中病毒的传播和控制及防御产生了重大的影响[23]。在本研究中,将网民评论词语作为节点,词语间的关联关系作为边,构建基于热点事件下的子事件演化网络。

网民话题讨论内容随着事件的发酵不断发展演进,使得评论词语基于彼此之间的关联关系逐渐形成代表网民意见的内容,因此,对于评论网络而言,发现评论网络中的不同意见群落以及其中的情感标签将有助于揭示网民意见与情感的演变。将网络划分为群落的依据大体上分为图划分和社群发现两类算法。其中图划分的主要特点是指定了群落规模和数量,目的在于把网络划分成更小、更容易管理的碎片从而进行研究;而社群发现主要是依据网络本身的拓扑结构来决定群落的数量和规模,即自然地识别网络内部不同群落。

本研究目的在于探索基于事件演化下网民意见和情感的演变过程,事先并不确定网民意见群落的具体情况,因此,本研究采用社群发现的 Louvain[22]算法对评论网络中的意见群落进行划分与识别。M. L. Wallace 等[24]也曾指出社群发现算法有助于规模较大的复杂网络的分析研究,并在主题的识别上具有明显的天然优势。

Louvain 算法是基于模块度的社群发现算法,该算法不需事先确定群落信息,并且在研究大型网络上具有较好的效率和效果表现。该算法通过不断将节点进行凝聚,实现将整个网络分解为多个群落。Louvain 算法的主要步骤如下:

(1)将网络中每个节点看成是一个独立的群落,即群落数目与节点个数相同;

(2)对每个节点 i,依次尝试把节点 i 分配到每个

邻居节点所在的群落中，计算分配前后的模块度变化 ΔQ，并记录 ΔQ 值最大的邻居节点，如果 MaxΔQ > 0，则把节点 i 分配到该邻居节点所在的群落，否则保持不变；

（3）对网络进行压缩，将所有同在一个群落内的节点压缩成一个新的节点，重复步骤（1）的过程，直至整个网络的模块度不再发生变化，即达到最大值。

为了验证网络中群落被划分得是否合适，研究中采用2004年 M. E. J. Newman 等[25]提出的模块度指标来对群落划分质量进行衡量，用 Q 表示。模块度的基本思想就是把划分出群落后的网络与相应的零模型进行比较，以度量群落划分的质量。而所谓的零模型就是指与所研究的网络具有某些相同的性质，而在其他方面完全随机的随机网络模型。目前在分析网络群落结构时，通常把待研究的网络和与其具有相同度序列的随机网络（一阶零模型）作比较[26]。因此，常用的模块度定义为：

$$Q = \frac{1}{2M} \sum_{ij} \left(a_{ij} - \frac{k_i k_j}{2M} \right) \delta(C_i, C_j) \qquad 式(1)$$

公式（1）中 M 是网络中的边数，A = (aij) 是网络的邻接矩阵，k_i 和 k_j 分别是网络中节点 i 和节点 j 的度值，C_i 和 C_j 分别表示节点 i 与节点 j 在网络中所属的群落：如果两个节点属于同一个群落，δ 取值为 1，否则为 0。事实上，一个网络的模块度就是该网络的群落内部边数与相应的零模型的群落内部边数之差占整个网络边数 M 的比例。有研究发现，当 Q 值在 0.3 – 0.7 时表明网络中出现了较强的群落结构[27]。

综上所述，Louvain 算法通过不断聚集各评论词语节点来实现模块度值最优的结果，从而确定评论网络中群落的划分。基于上述理论和思想，本研究沿事件发展脉络来识别评论网络中的意见群落，动态地跟踪与分析意见群落的演化过程。

3.1.2　情感词典　本研究选用大连理工大学的中文情感词汇本体库[28]，其分类体系是在国外比较有影响的 Ekman 的 6 大类情感分类体系的基础上构建的，最终词汇本体中的情感共分为 7 大类。情感词汇本体中的词性种类一共分为 7 类，每个词在每一类情感下都对应了一个极性，其中，0 代表中性，1 代表褒义，2 代表贬义，3 代表兼有褒贬两性。其中，一个情感词可能对应多个情感，情感分类用于刻画情感词的主要情感分类，辅助情感为该情感词在具有主要情感分类的同时含有的其他情感分类。情感分类按照论文《情感词汇本体的构造》[28]所述，情感分为 7 大类、21 小类。情感强度分为 1、3、5、7、9 五档，9 表示强度最大，1 为强度最小。该词典共有情感词 27 466 个，其中褒义词 11 229个，贬义词 10 782 个，情感词汇本体格式如表 1 所示：

表 1　情感词汇本体格式（部分）

词语	词性种类	词义数	词义序号	情感分类	强度	极性
无所畏惧	成语	1	1	好	7	1
手头紧	成语	1	1	恶	7	0
周到	形容词	1	1	好	5	1
言过其实	成语	1	1	恶	5	2

情感词是否全面对情感分析具有重大意义，本文使用的情感词典包含基础词典和基于特定事件情境下的情感词典。本研究将大连理工大学情感词典作为基础词典，但在研究特定事件情景时单单使用基础词典中的情感词往往是不够的，在研究热点事件用户意见挖掘问题时，需要有关于此次热点事件的高频情境词汇。因此，结合人工筛选，参照大连理工大学对情感词的打分情况，建立针对特定事件情境下的领域情感词典[9]。由于研究的需要，本研究对一些词语进行了特殊处理，例如“中兴”在情感词典中为褒义词，带有情感极性，而在本研究中该词是具有问题代表性的名词，大量出现会影响研究结果，因此，将此类词语从情感词典中删除。同时也对评论内容中的否定词进行了特殊处理，观察情感词邻居是否有否定词存在，若情感词周边存在否定词，则将否定词与情感词合并为一个词语，如“无法理解”“不是好人”等，按照这种规则，参照情感词典对词语进行情感极性和分类的倒置，将此类词语增添在领域情感词典中，部分领域情感词典如表 2 所示：

表 2　领域情感词典（部分）

词语	词性	强度	极性	情感分类	词语	词性	强度	极性	情感分类
狗	名词	7	2	恶	台独	名词	7	2	恶
凉了	动词	5	2	恶	白眼狼	成语	7	2	恶
制裁	名词	7	2	怒	傻了吧唧	成语	5	2	恶

3.2　研究思路框架

网络舆情事件所固有的传播速度快、不易控制等特点，使得舆情事件的引导控制工作成为政府、社会各界关注的重点，动态地跟踪舆情事件演进过程中网民的话题讨论内容以及情感变化，可以整体把握热点事件的宏观发展方向。

研究中，首先对所获取事件的网民评论进行数据预处理，去除重复、垃圾评论，将评论数据进行分词处理，进而构建评论词网络，使用情感词典赋予网络中情

感词节点相应的情感分类,采用 Louvain 算法划分网民评论内容群落。从网民话题评论内容以及情感分析两个角度对评论数据进行分析,通过网民话题讨论内容挖掘舆情事件整体演化过程中的网民意见,依据网民情感波动情况反映网民对于整个事件的态度,最后采用局部视角,将评论内容分析和情感分析结合,进一步揭示同一话题随着事件演进的不同时期网民讨论内容以及网民情感的变化情况,再结合评论内容解释网民情感变化原因。研究思路见图 1。

4 网民评论网络特性分析

4.1 数据选取与预处理

新浪微博是由新浪网推出,提供微型博客服务类的社交网站,它是一个为大众提供娱乐休闲生活的信息分享和交流平台。据央视财经统计,截止到 2017 年第四季度,微博活跃用户 1.72 亿,月活跃用户 3.92 亿。因此,选取新浪微博作为本研究数据源,选择政府、社会、民众都比较关心的"中美间贸易摩擦"为研究话题,获取 2018 年 3 月 23 日中美贸易战正式打响开始至 2018 年 4 月 30 日"中美间贸易摩擦"事件下子事件中热门博文的网民评论,共 3 313 条,选取热点事件子事件、博文以及评论数见表 3。

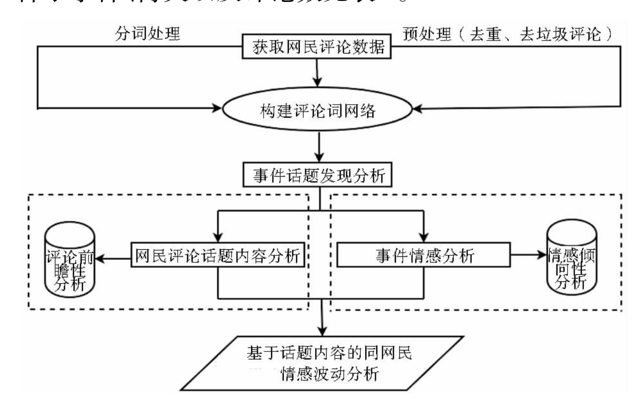


图 1 研究思路框架

表 3 中美间贸易摩擦节点事件、博文及评论数

序号	中美间贸易摩擦子事件	时间	博文	评论数
T1	中美贸易战开始标志	3.23	特朗普签署对华关税备忘录 表示这只是开始	891
T2	中国对于贸易战的反击	4.2	中美贸易摩擦升级！中国为何选在 4 月 1 日半夜重拳反击?	345
T3	美国对中兴实施一系列制裁行动	4.17	美国封杀中兴,一颗芯片都不卖,直到 2025 年！中美科技战启幕	610
T4		4.19	中心被制裁最新进展:已停止相关业务 向美申请解释	578
T5		4.28	台当局宣布将中兴列入出口管制对象	889

使用正则表达式对评论内容数据进行清洗,如@、数字、英文字母、用户名、表情等,只提取其中的主题评论内容,再将清洗后的评论数据进行分词处理,分词方法采用 JIEBA 分词,逐条进行分词,去掉停用词、标点等无意义的词,用经过清洗和分词后的评论词构建基于事件演化的评论词网络,在网络中,使用 R 语言将情感词典中 7 类情感分类与各个情感词节点进行匹配,赋予情感词节点属性。基于事件 T1 所构建评论词网络见图 2。

各个子事件网络整体节点个数、边数与最大联通子图个数、边数如表 4 所示,由于子图间规模相差悬殊,本研究提取网络中的最大联通子图进行相关分析研究。

表 4 评论网络基本指标

子事件	T1	T2	T3	T4	T5
整体网络节点个数	2 698	1 240	2 244	2 603	2 855
最大联通子图节点个数	2 518	1 169	2 145	2 526	2 725
整体网络节点连接边数	25 318	11 049	24 948	32 782	27 012
最大联通子图节点连接边数	25 180	10 958	24 859	32 703	26 896



图 2 T1 子事件评论词网络

4.2 社群规模分析

本研究采用社群发现算法中的 Louvain 算法,根据评论网络的拓扑结构来对话题社群进行识别和分类,基于事件演化过程动态探索领域话题群落的演变过程。由于各子事件评论网络拓扑结构的不同,导致不同子事件中所划分出的群落数目与规模有所差异。T1-T5 这 5 个事件窗口下各个评论网络中知识群落的

数量及规模如表 5 所示。其中第一行代表各事件窗口,第一列代表了群落编号。为了清晰地展示各事件窗口中的群落情况,表 5 中每一事件窗口下的话题群落根据规模由高到低进行排序。

表 5 子事件各个群落规模

群落	T1	T2	T3	T4	T5
群落 1	315	186	341	373	389
群落 2	196	182	271	234	368
群落 3	190	103	258	227	182
群落 4	160	95	227	199	178
群落 5	153	95	159	157	160
群落 6	152	87	124	144	148
群落 7	138	57	120	137	138
群落 8	132	55	111	128	128
群落 9	131	47	110	121	125
群落 10	116	40	96	119	123
群落 11	114	39	77	117	119
群落 12	100	34	75	112	100
群落 13	94	33	36	82	86
群落 14	93	31	33	81	75
群落 15	81	30	23	75	67
群落 16	71	25	20	64	63
群落 17	62	9	13	58	58
群落 18	57	7	6	47	52
群落 19	54	5		23	45
群落 20	49	5		20	38
群落 21	43	4		9	38
群落 22	10				30
群落 23	7				5
群落 24					5
群落 25					5

表 5 中对各事件窗口中知识群落的数量以及规模进行了统计汇总。为了考察所划分出的知识群落是否合理,分别计算各个事件窗口下的评论网络的模块度,如表 6 所示:

表 6 基于事件的评论网络模块度

子事件	T1	T2	T3	T4	T5
模块度	0.553	0.563	0.522	0.502	0.609

根据表 6 可知,各个事件窗口下的评论网络的模块度值均在 0.3 – 0.7 之间,说明研究中利用 Louvain 算法所划分的话题群落较为合理。子事件评论网络中社群数量代表子事件话题讨论的丰富程度,由表 5 可知,随着热点事件的演化,网民对于各子事件引发的话题讨论数量存在着不同程度的差异,其中,T1 事件(“特朗普签署对华关税备忘录 表示这只是开始”)与 T5 事件(“台当局宣布将中兴列入出口管制对象”)的发生引发了网民较大程度的关注与讨论。

4.3 变异系数分析

为了比较热点事件演变过程中各个子事件的话题讨论集中程度,本研究采用变异系数这一统计指标来探索基于事件演化过程中不同规模的子事件群落之间的话题讨论集中程度。该指标消除了不同社群规模和关联数量的影响。

$$C_v = \frac{\sigma}{\mu} \qquad 式(2)$$

公式(2)中,C_v 表示变异系数,σ 代表话题群落规模的标准差,μ 是群落规模的平均数。C_v 描述了话题群落标准差与平均数之比。变异系数水平越高说明群落越离散,话题讨论越集中。各个子事件变异系数如表 7 所示:

表 7 各个子事件变异系数

子事件	T1	T2	T3	T4	T5
变异系数	0.627	0.946	0.915	0.711	0.890

分析表 7 中数据可知,各个子事件变异系数整体上处于波动状态,基于事件演化过程中话题讨论的集中程度有所不同,T1 事件的变异系数最低,随着事件发展,变异系数有所上升,但仍处于波动变化状态,这一现象从事件演化角度揭示了在舆情事件爆发的初期,网民话题讨论较为分散,没有形成高度集中的主题,随着舆情事件的演化,关键节点事件逐渐出现,将网民讨论话题向着统一的方向引流。

5 中美贸易战话题发现及情感波动分析

5.1 中美间贸易摩擦事件话题发现分析

5.1.1 网民评论话题内容分析 为了从宏观层面上了解 T1 – T5 整个事件发展的全过程,笔者选取每个子事件下评论网络中规模较大的社群,网络社群中高点度的节点可以体现网民讨论的主要话题内容。由于篇幅有限,选取子事件评论网络中各个群落中度数较高的词语节点,如图 3 所示,对 5 个子事件的话题内容进行分析,可以发现网民对于一些话题进行了较为持续地继承讨论,如贸易、关税等话题。随着舆情事件的发酵,会涌现出更多新话题的讨论,如 T2 事件的农业、经济、电脑系统等;T3 事件中的芯片、科技、研发、核武器等话题;T4 事件中的科研人才、科学技术等话题;T5 事件中的台独、联发科、发展等话题。值得注意的是,“中美贸易战”话题社群在前 4 个子事件中都是排名第 1 位或第 2 位的大型社群,但在 T5 事件(台当局宣布将

中兴列入出口管制对象）中，网民评论内容中包含“中美贸易战”的话题社群只排在第3位，网民的主要讨论内容集中在反对台独、大陆与台湾发展问题等话题，削弱了中美贸易战的相关话题讨论，这个现象说明在热点事件的演化中逐渐催生出了关于事件的衍生话题。

T1	社群1	美国	中国	关税	贸易战	世界	保护	产业	车	民族	欧洲
	社群2	发展	汽车	进口	税	川普	工业	税收	品牌	进口车	农业
	社群3	特朗普	总统	中美	政治	国是	威胁	战略	下台	精英	社会
	社群4	手机	台湾	苹果	兄弟	贸易逆差	平民	高通	华为	领导人	芯片
T2	社群1	中国	美国	贸易战	产品	农业	经济	加税	制裁	影响	发展
	社群2	苹果	国家	企业	富士康	手机	市场	生产	抵制	代工	国产手机
	社群3	关税	进口	销售	电子	不卖	死	担心	美国佬	喉咙	困难
	社群4	爱国	安卓	微软	发明	系统	电脑	耐克	媒体	反美	谷歌
T3	社群1	芯片	手机	华为	生产	高通	处理器	麒麟	半导体	核心	世界
	社群2	中国	国家	研发	科技	自主	核武器	发展	核心技术	基础	国外
	社群3	美国	中兴	制裁	员工	处罚	违反	罚	实力	商务部	软件
	社群4	企业	公司	行业	技术	国内	政府	高科技	人才	创新	研究
T4	社群1	中国	美国	贸易战	生活	日本	落后	强大	祖国	民族	产品
	社群2	国家	中兴	华为	协议	自信	员工	新闻	影响	和解	奖金
	社群3	研发	人才	经济	国内	国人	移民	高端	条件	社会	努力
	社群4	芯片	投入	行业	房地产	核心	资金	高铁	投资	战争	计算机
T5	社群1	台湾	大陆	发展	台独	中兴	经济	日本	民进党	支持	制裁
	社群2	美国	狗	国家	蔡英文	关系	走狗	航母	教训	土地	对付
	社群3	中国	统一	市场	世界	联发科	贸易战	手机	中美	资本	厂家
	社群4	敌对	研究	毛线	芯片	韩国	房地产	中华民族	加戏	建筑工人	技术含里

图3　事件群落中度数中心度较高词语节点截图（部分）

5.1.2　网民评论话题前瞻性分析　为了更细致具体地揭示网民评论话题内容，笔者从微观层面对网民话题讨论内容进行聚焦，通过分析事件演进过程中网民的话题社群内容，如图4所示，可知T1子事件（2018年3月23日中美贸易战初始）中，网民评论网络中关于芯片、华为手机的话题社群就已经出现且度数中心度较高。根据表3回溯舆情事件的发展脉络，3月23日时间节点中，是关于特朗普签署对华关税备忘录这一事件，T3事件（美国制裁中兴）发生日期为近一个月后的4月19日，子事件话题群落如图5所示，这一现象体现了网民对于此次舆情事件演化的前瞻性，对于事件的思考预判达到了一定的深度和层次，预测到了事件发展的走向。在分析过程中，笔者发现这一现象并不是特例，例如，在T3事件（中兴被制裁）中，芯片词语节点与华为词语节点高度共现。针对这一现象，政府以及相关部门应予以高度重视，充分发挥微博平台的民意收集作用，对于网民讨论话题内容可以进行提取，进而充分利用。

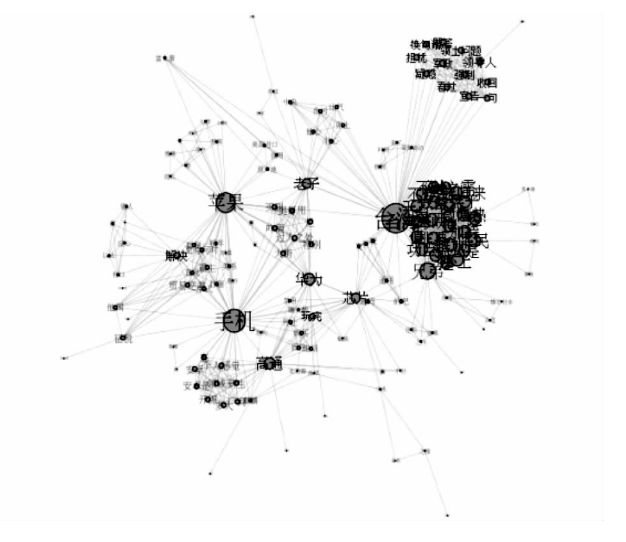


图4　T1事件子群落



图5　T3事件子群落

5.2　网民情感倾向性分析

通过分析评论的倾向性可以从宏观层面上把握网民在整个事件演化过程中的情绪波动状况，本研究对评论内容情感以乐、好、怒、哀、惧、恶、惊七大类情感的变化进行追踪，由于评论网络中的每一个情感词节点都根据情感词典被分别赋予上述七大类属性，因此笔者使用R语言对子事件T1-T5评论网络中情感词节点进行统计分析，可以得到网民对于中美贸易间摩擦话题整体情感倾向性情况，网民情感倾向性变化见图6。

	T1	T2	T3	T4	T5
恶	42.90%	43.31%	39.38%	42.82%	46.01%
怒	1.09%	1.27%	2.50%	2.87%	0.70%
乐	7.92%	8.28%	7.50%	7.57%	7.88%
惧	4.92%	3.18%	6.56%	3.13%	3.76%
惊	1.09%	0.64%	0.63%	0.26%	0.70%
好	37.43%	37.58%	36.25%	36.29%	34.98%
哀	4.64%	5.73%	7.19%	7.05%	5.37%

图 6　网民情感倾向

从宏观层面上分析，在事件演化过程中，网民对于此次事件“恶”“怒”“哀”的情绪是增加的，“好”“惊”“乐”“惧”的情感在总体上是削减的，体现了网民随着事件的发展演变消极情绪逐渐增多，积极情绪逐渐减少。网民对于中美间贸易摩擦话题情感向两种极端情感集中，“恶”和“好”是表达网民情感的主力军，说明网民相对客观地对待此次事件，认为此次事件既是机遇也是挑战。值得一提的是，在 T2 事件向 T3 事件发展过程中，“惧”的情感显著增加，T3 事件向 T4 事件发展过程中，“惊”的情绪显著增加而“怒”的情绪显著减少。回溯整个事件发展过程，从中美贸易战正式打响到中国予以反击到中兴被制裁再到台湾涉足此次事件，网民对于中兴被制裁、美国遏制我国科技发展步伐感到惊恐，在大陆台湾历史长期遗留问题的政治背景下，网民对于台湾涉足此次事件并没有愤怒，更多地是惊奇和指责。

5.3　基于话题内容的网民情感波动分析

在舆情事件的持续演化过程中，网民对于相同话题的情感会随着事件的发展产生不同的情感倾向，网民的情感将呈现一定程度的波动。为了追踪网民情感的细微变化，笔者从局部角度选取了不同子事件 T1、T4 事件中相同讨论话题的社群，如图 7、图 8 所示：

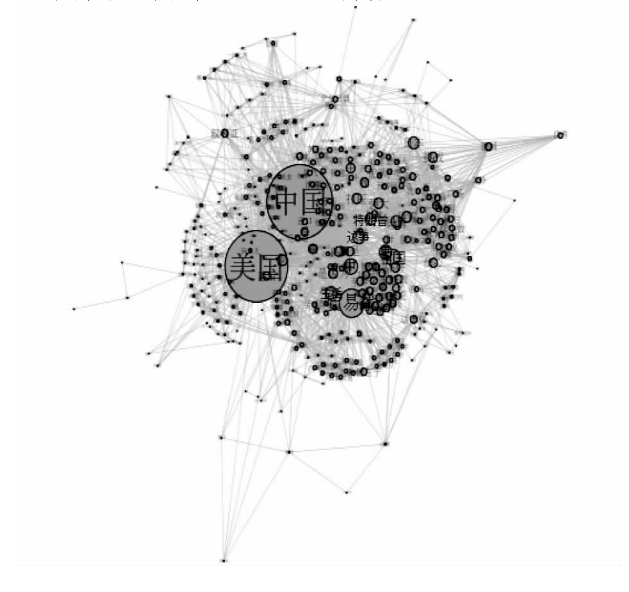


图 7　T1 子事件中美贸易话题社群

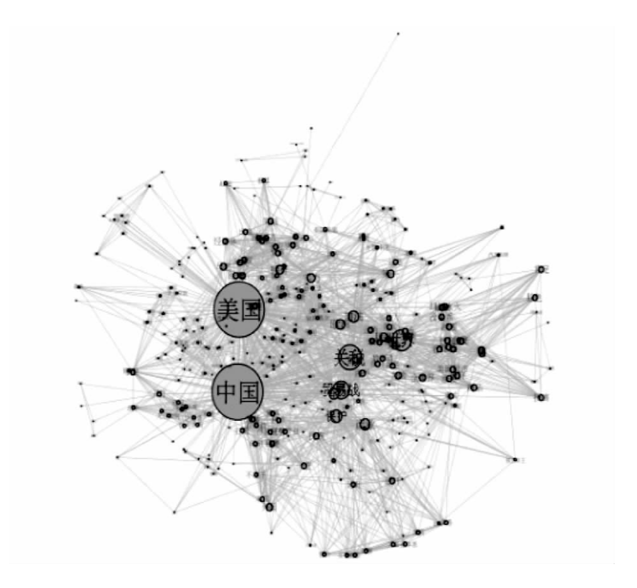


图 8　T4 子事件中美贸易话题社群

由图 7 与图 8 中可以看出两个社群话题都主要是关于中美贸易相关问题的讨论，两个子事件下中美贸易话题网民情感波动变化情况如图 9 所示：

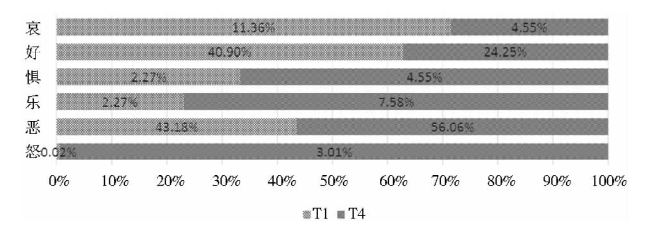


图 9　T1、T4 子事件中美贸易话题网民情感波动分析

根据图 9 中的情感变化对比可以发现，事件由 T1（中美正式开战）演化到 T4（美国制裁中兴）过程中，网民对于此事件“恶”“惧”“乐”“怒”的情感增多了，“好”“哀”的情感减少了，为深入探究情感变化原因，本文结合群落话题内容来理解网民的情感变化。T1、T4 事件中美贸易社群中除中国、美国、贸易战节点外高点度节点如表 8 所示：

表 8　社群高频节点度数

T1		T3	
节点	度数	节点	度数
关税	282	中兴	554
保护	166	芯片	525
产业	133	研发	260
车	130	华为	231
股市	69	投入	209

由表 8 可知，T1 事件中美贸易社群话题主要侧重于关税、产业、股市等方面，T4 事件中美贸易社群话题主要关于中兴、华为、芯片、研发等，根据社群讨论内容以及情感的波动变化，可以总结出网民的观点和态度。

美国对于中兴的制裁行为,加重了网民的对于国家未来科技发展的恐慌以及愤怒、憎恶、指责的情绪,与此同时,网民“乐”的情感增多,“哀”的情感减少了,体现了网民对于民族产业、自主创新研发的信心和决心。

6 结语

(1)评论网络群落发现以及变异系数可以有效地衡量网民话题讨论的规模与集中程度,网民话题讨论集中程度受到子事件的影响,与热点事件演化的不同阶段有关。

(2)舆论衍生话题对事件网络舆情发展有持续性影响。随着网民讨论话题广泛性的升高,逐渐出现了舆论衍生话题,舆论衍生话题对于舆情事件的发展方向具有很大的影响。

(3)网民对于事件的演化具有一定程度上的前瞻性。网民对于舆情事件发展的准确预判并非偶发现象,舆情的意见收集汇总在某种程度上形成了一种群体智慧[29],在社会学领域中充分利用群体意见对舆情问题的管理具有重要意义。

(4)评论网络中赋予情感词节点情感分类属性可以体现事件演化过程中网民的情感变化。网民对于事件的情感变化多数受媒体报道、时事新闻渠道、时事新闻内容影响,因此政府和相关部门充分利用新闻媒体,把握好网民情感变化节点,有针对性地进行引导管控,对于舆情的导向具有重要意义。

本文基于复杂网络分析思想,采用 Louvain 算法来识别划分子事件评论网络中的话题群落,基于情感词节点属性确定事件演化过程中网民的情感变化。但研究工作也还存在有待完善的地方,所选取的新浪微博数据源,数据覆盖面不够广,在后续的研究中将选取更广泛的话题以及数据源,对此类问题做更为深入的研究。

参考文献:

[1] 刘毅. 网络舆情研究概论[M]. 天津:天津人民出版社,2007.

[2] CHEN Y, AMIRI H, LI Z, et al. Emerging topic detection for organizations from microblogs[C] // Proceedings of the 36th international ACM SIGIR conference on research and development in information retrieval. Dublin: ACM, 2013: 43－52.

[3] STILO G, VELARDI P. Efficient temporal mining of micro-blog texts and its application to event discovery[J]. Data mining and knowledge discovery, 2016, 30(2): 372－402.

[4] HU M, LIU B. Mining and summarizing customer reviews[C] // Proceedings of the ACM SIGKDD international conference on knowledge discovery & data mining. Seattle: ACM, 2004, 168－197.

[5] 李亚星,王兆凯,冯旭鹏,等. 基于实时共现网络的微博话题发现[J]. 计算机应用,2016,36(5):1302－1306.

[6] 宋丽娜,冯旭鹏,刘利军. 基于 SOM 聚类的微博话题发现[J]. 计算机应用研究,2018,35(3):671－679.

[7] 唐晓波,王洪艳. 基于潜在语义分析的微博主题挖掘模型研究[J]. 图书情报工作,2012,56(24):114－119.

[8] 唐晓波,肖璐. 基于依存句法分析的微博主题挖掘模型研究[J]. 情报科学,2015,33(9):61－65.

[9] 王正成,袁竹星. 面向主题的微博意见领袖挖掘研究[J]. 情报科学,2018,36(3):112－116.

[10] 肖璐,唐晓波. 基于句子成分的微博热点主题挖掘模型研究[J]. 情报科学,2015,32(11):44－47.

[11] 赵妍妍,秦兵,刘挺. 文本情感分析[J]. 软件学报,2010,21(8):1834 －1848.

[12] PANG B, LEE L, VAITHYANATHAN S. Thumbs up? sentiment classification using machine learning techniques[EB/OL]. [2018－05－08]. https://arxiv. Org/pdf/cs/ 0205070. pdf.

[13] PANG B, LEE L. A sentimental education: sentiment analysis using subjective summarization based on minimum based on minimum cuts[EB/OL]. [2018－05－08]. https:// arxiv. org/pdf/cs/0205070. pdf.

[14] BOLLEN J, MAO H, PEPE A. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena[J]. Computer science, 2009, 44(12): 2365－2370.

[15] ZOU X, YANG J, ZHANG J. Microblog sentiment analysis using social and topic context[J]. PLoS one, 2018, 13(2): e0191163.

[16] ZHANG L, GHOSH R, DEKHIL M, et al. Combining lexicon-based and learning-based methods for Twitter sentiment analysis[EB/OL]. [2018－05－10]. http://www. hpl. hp. Com / techreports / 2011 / HPL-2011-89. pdf.

[17] 唐晓波,兰玉婷. 基于特征本体的微博产品评论情感分析[J]. 图书情报工作,2016,60(16):121－127.

[18] 杨亮,林原,林鸿飞. 基于情感分布的微博热点事件发现[J]. 中文信息学报,2012,26(1):84－90.

[19] 黄卫东,陈凌云,吴美蓉. 网络舆情话题情感演化研究[J]. 情报杂志,2014,33(1):102－107.

[20] 张向阳,那日萨,孙娜. 基于有向网络的在线评论情感倾向性分析[J]. 情报科学,2016,34(11):66－69.

[21] 杨锋,彭勤科,徐涛. 基于随机网络的在线评论情绪倾向性分类[J]. 自动化学报,2010,36(6):837－844.

[22] BLONDEL V D, GUILLAUME J L, LAMBIOTTE R, et al. Fast unfolding of communities in large networks[J]. Journal of statistical mechanics, 2008, 2008(10): 155－168.

[23] LEWIS T G. 网络科学:原理与应用[M]. 陈向阳,巨修炼,等译. 北京:机械工业出版社,2011:1－12.

[24] WALLACE M L, GINGRAS Y, DUHON R. A new approach for detecting scientific specialties from raw cocitation networks[J]. Association for Information Science and Technology, 2009, 60

(2): 240 – 246.
[25] NEWMAN M E J, GIRVAN M. Finding and evaluating community structure in networks[J]. Physical review E, 2004, 69(2): 1 – 15.
[26] 汪小帆,李翔,陈关荣.网络科学导论[M].北京:高等教育出版社,2012:130 – 140.
[27] TANG L,LIU H.社会计算:社区发现和社会媒体挖掘[M].文益民,闭应洲, 译. 北京: 机械工业出版社, 2013: 20 – 30.
[28] 徐琳宏,林鸿飞,潘宇,等.情感词汇本体的构造[J]. 情报学报,2008, 27(2): 180 – 185.
[29] WECHSLER D. Die messung der intelligenz erwachsener[M]. Bem-Stuttgart: Huber, 1964.

作者贡献说明:

张海涛:提出研究思路与方法、进行数据分析及论文修订;

刘雅姝:负责数据采集与分析处理、论文初稿撰写;

张枭慧:负责数据收集、整理;

宋拓:负责论文修订。

Research on Topic Discovery Based on Modularity and Sentiment Fluctuation of Internet Users
——Taking Sina Weibo's "China-US Trade Friction" as an Example

Zhang Haitao[1,2]　Liu Yashu[1]　Zhang Xiaohui[1]　Song Tong[1]
[1] Management School of Jilin University, Changchun 130022
[2] The Information Resource Research Center of Jilin University, Changchun 130022

Abstract: [**Purpose/significance**] Exploring topical communities and sentiment fluctuations of Internet users and grasping the process of development of events have great significance to control the development direction of the events and lead guidance of the network public opinion in the new period. [**Method/process**] Based on the theory of complex networks, the study constructed sub event network based on co-occurrence relations among user comments, identifying topic community in sub-event commenting networks through community discovery algorithms and giving the attribute to emotion word according to the emotional dictionary. The study dynamically tracked the opinions and emotions of Internet users based on the evolution process of events. [**Result/conclusion**] The conclusion showed that the commenting network community discovery and coefficient of variation method can effectively measure the scale and concentration of Internet users' topic discussion; emotional word sentiment classification attribute method can reflect the emotional changes of Internet users in the process of event evolution; the derived topic of public opinion has a continuous influence on the development of the event public opinion; the content of the topic discussion of Internet users has some foresight to the evolution of the event.

Keywords: complex network　community discovery　topic discovery　sentiment analysis

下　期　要　目